**MEMORANDUM**

From: Glenn Rice

To: Dr. Craig

Date: 5/2/2022

Subj: Analysis of Gas Prices in the United States

**Executive Summary**

Gasoline prices play a huge part in our everyday lives. It is how we get to work, school, doctor’s appointments, and just about anything else you can think of. When you drive down the road, you will most likely see multiple massive signs that show gas and diesel prices in vibrant colors, and they are always changing. Gasoline is a substitute good, meaning there is not much brand loyalty when it comes to getting gas. If you notice that Gas Station A is twenty-five cents cheaper than Gas Station B, you are most likely going to get gas at the cheaper place.

I am interested in gas prices because it is such a relevant topic in almost everyone’s lives, and they are changing so much. I wanted to look at a few different factors to gain knowledge as to why gas prices are what they are. The goal of this paper is to try to better understand why gas prices are listed as they are. In this study, we will look at multiple areas of gasoline prices including crude oil, prices relative to region, and annual gasoline consumption by region.

**Statement of Hypotheses**

While looking at data from the United States Energy Information Administration, it was very intriguing to see crude oil prices, retail prices, annual barrel consumption and how it fluctuates over any specified time period, which could be weeks, months, or years. All of my data are complied from a series of four different datasets from the United States Energy Information Administration. The first one being the nation’s monthly prices of regular grade gasoline and crude oil from January 2008 to December 2020. The second dataset that I will use is the United States’ average weekly gas price by region from August 20, 1990 to April 25, 2022. Third, I will look at the United States’ annual barrels of oil consumption by region. Finally, although I do not run a statistical test due to the limited data I found on state and federal tax rate, I will include it in a later section due to its relevance. I will test the subsequent hypotheses using Minitab.

***Hypotheses:***

Test 1

There is no formal hypothesis for my first test, which will be a regression, but I think this is the most appropriate place to introduce it. I will determine how much of a factor crude oil is to the price of regular gasoline in the Unites States.

Test 2

Null Hypothesis: All of the nation’s weekly regional gas prices from August, 1990-2022 are equal.

Alternative Hypothesis: Not all of the nation’s weekly regional gas prices from 1990-2022 are equal.

Test 3

Null Hypothesis: All of the nation’s regional annual barrel consumption from 1981-2021 are equal.

Alternative Hypothesis: Not all of the nation’s regional annual barrel consumption from 1981-2021 are equal.

**Data & Analysis Methodology**

The data I gathered was from various website extensions of the United States Energy Information Agency. While all of my data came from one government agency’s website, it came from different observational studies, making the overall time periods and unit of measure unique to each dataset. The first dataset that I attained was the national average of regular gas prices each month from 2008 to 2020. The second dataset that I put together was the weekly average of gas prices according to PADD from 1990-2022. PADD stands for Petroleum Administration for Defense Districts. The United States is split into 5 regions: East Coast, Midwest, Gulf Coast, Rocky Mountains, and West Coast. In this paper, if I refer to PADD or region, I am using them as interchangeable terms. The third dataset that I found was the annual regional barrel consumption by the United States. It is measured in thousands in the dataset, so any number that you see directly off of Minitab should be multiplied by 1000 to get the real result. Finally, I collected the tax rates at the state and federal level for each individual state. I will not be running a test on this because I could only find the rates for 2021, but I thought it would be negligent to just ignore them completely. Note: “regular” means the grade of gasoline. Regular grade gasoline is what you’ll see on the pump as 87 octane.

To test the hypotheses that were mentioned earlier, I will first use a scatterplot with regression, with the price of regular gas on the Y-Axis and the price of crude oil on the X-Axis to determine the R-Squared value, to see how much of the price of regular gas is explained through its relationship with the price of crude oil. The data that I will be using is the national average collected monthly beginning in January of 2008, all the way up to December of 2020. The second test I will be conducting is an ANOVA one-way test supplemented with a Tukey comparison. This will help me determine if the five regions in the United States have statistically similar or dissimilar data when looking at the regions weekly average regular gasoline price from 1990 to 2022. Finally, I will use another ANOVA one-way test with a Tukey comparison to look at regional consumption of barrels of oil. The goal is to determine if certain regions consume different amounts. The data I will be using is measured in thousands and are from the year 1981 to 2021. To conduct these tests, I will use Minitab for all of them. Some of my data directly translated from my sources to Minitab, while others had to be converted from Excel.

**Results**

***Regression Analysis: Scatterplot (Test 1)***

The first analysis I ran aimed to see how closely related regular gas prices and crude oil prices were. I ran a scatterplot, with regression, and put crude oil on the X-Axis and gas prices on the Y-Axis. I predicted that there was going to be a strong correlation, but I did not think it was going to be as strong as it actually was. My findings showed that there was a R-Squared value of 93.3%, meaning that about 93% of the variability is gas prices can be explained with the price of crude oil. Chart, scatter chart

Description automatically generatedThe regression equation was also calculated, which could help predict where a new data point would most likely land on the scatterplot.

The regression equation is: Regular Gas = 1.215 + 0.9530 Crude Oil

After looking at the graph and regression equation, it is safe to say that crude oil price and regular gas prices are positively correlated.

***One-Way ANOVA (Test 2)***

The hypothesis that I tested with this method was that all of the nation’s weekly regional gas prices from August, 1990-2022 are equal. The test showed that there was enough evidence to conclude that not all of the groupings were statistically similar. Since the P-Value was less than .05, it is not statistically possible for the five different regions to have the same gas prices over the tested time period. Therefore we must rejejct the null hypothesis. This means the alternative hypothesis is correct, which is that all of the nation’s weekly regional gas prices from August, 1990-2022 are not equal. It should be noted that this is with a confidence interval of 95%. The higher the confidence interval is, the broader the interval must be, so it has a higher probability of being correct.

Text

Description automatically generated with medium confidence

Because I used a Tukey comparison, I was able to look at how the five different regions were able to fit into groupings. The West Coast was in a group of its own with the highest price. The Rocky Mountains and East Coast fit into the second group with the second highest groupings of means. Midwest was unique because it fit into both group B and group C. And the Gulf Coast had the lowest average mean, placing it in group C. The Tukey groupings are based off of the interval plot which is included below. The groupings are determined by if the confidence intervals are overlapping or not.

Table

Description automatically generated

Chart, line chart

Description automatically generated

***One-Way ANOVA (Test 3)***

The last test that I ran tested the hypothesis that all of the nation’s regional annual barrel consumption from 1981-2021 are equal. It compared the means of consumption, separated by regions, dating back to 1981. When examining the P-Value of the one-way ANOVA test that I ran, it came up with a P-Value of 0.000, which means there is no possible way that the five regions can be statistically similar when it comes to oil consumption. The P-Value is less than .05 meaning we must reject the null hypothesis. From rejecting the null hypothesis, we must accept the alternative hypothesis, which was all of the nation’s regional annual barrel consumption from 1981-2021 are not equal.

Text

Description automatically generated with low confidence

I also used the Tukey comparison again and saw that there were only four different groupings for the five regions. The East Coast (Group A) consumed the most. The Gulf Coast and the Midwest was in group B. The West Coast was in group C, and the Rocky Mountains was in group D. I thought that it was surprising that the West Coast did not consume the most or second most; I did not have much of a reason to believe that, that was just a guess out of sheer prior knowledge. When looking at the interval plot, the overlapping confidence intervals are where the groupings are placed, and you can see that the Midwest and the Gulf Coast have overlapping confidence intervals.

Table

Description automatically generated

Chart

Description automatically generated

**Discussion and Conclusion**

Gasoline prices in America are always changing. In this Analysis, we looked at the prices of regular gas and crude oil, prices in regard to region, and consumption within each region. When using a scatterplot with a regression line added, we found that 93.3% of the variability in gas prices can be accounted for by crude oil prices. Through both of our ANOVA tests, we found that prices and consumption varied considerably when looking at the five regions. When looking at the Tukey comparisons between test 2 and test 3, I thought it was interesting that they did not follow an order when it came to the higher mean/grouping. The West Coast had the highest prices in group A but was the fourth lowest in group C when we looked at consumption. This might also have to do with the tax rates that I mentioned earlier. The following table shows the tax rate for each region in 2021.

Table

Description automatically generated with medium confidence

If I was able to obtain more years for state tax rate, I would have liked to run a multiple regression to determine how much more variability I could’ve accounted for. Overall, gasoline prices in the United States have many components to it, and this analysis got a good hold on a lot of the factors that play into the prices.

Sources

U.S average monthly gasoline and crude oil prices, 2008-2020

https://www.eia.gov/energyexplained/gasoline/price-fluctuations.php

Weekly Gasoline and Diesel Retail Prices

https://www.eia.gov/dnav/pet/pet\_pri\_gnd\_dcus\_nus\_w.htm

Crude Oil Production

https://www.eia.gov/dnav/pet/PET\_CRD\_CRPDN\_ADC\_MBBLPD\_A.htm

State Gasoline and Diesel Fuel Taxes

https://www.eia.gov/todayinenergy/detail.php?id=47376#